Economics 628 **Term 1 2017/2018**

Topics in Applied Econometrics I **Hiro Kasahara**

Homework 5

(Due: Wednesday, November 15 at the start of the class)

Note: Study groups *discussing* the problems are strongly encouraged. But please write your own answers and submit your own programs (no copy and paste from your classmate's program!).

1 OLS, IV, and LATE

Suppose that a firm makes a decision to import by profit maximization:

$$\begin{aligned} d &= 1\left[\pi(1,\epsilon) - \pi(0,\epsilon) - C(z,\xi) > 0\right], \\ \pi(d,\epsilon) &= \exp(\alpha + \beta d + \epsilon) = \exp(\alpha + (\mu_\beta + \eta)d + \epsilon), \text{ where } \beta = \mu_\beta + \eta, \\ C(z,\xi) &= \exp(-c_0 - c_1 z - \xi), \end{aligned}$$

where $d = 1$ if a firm imports and $d = 0$ otherwise. $\pi(d,\epsilon)$ is a profit when a firm makes an import decision $d \in \{0,1\}$, where ϵ is an unobserved profit shock (e.g., productivity shock). $C(z,\xi)$ is a fixed cost of importing, where z is an observed cost shifter and ξ is an unobserved cost shock.

We assume that (ξ,ϵ,η) is randomly drawn from a jointly normal distribution so that

$$\begin{pmatrix} \eta \\ \epsilon \\ \xi \end{pmatrix} \overset{iid}{\sim} N(0,\Sigma), \quad \text{where } \Sigma = \begin{pmatrix} \sigma_\eta^2 & \rho_{\eta\epsilon}\sigma_\eta\sigma_\epsilon & \rho_{\eta\xi}\sigma_\eta\sigma_\xi \\ \rho_{\eta\epsilon}\sigma_\eta\sigma_\epsilon & \sigma_\epsilon^2 & \rho_{\epsilon\xi}\sigma_\epsilon\sigma_\xi \\ \rho_{\eta\xi}\sigma_\eta\sigma_\xi & \rho_{\epsilon\xi}\sigma_\epsilon\sigma_\xi & \sigma_\xi^2 \end{pmatrix}.$$

Then, the model can be rewritten as:

$$\begin{aligned} \ln \pi_i &= \alpha + (\mu_\beta + \eta_i)d_i + \epsilon_i, \\ d_i &= \begin{cases} 0 & \text{if } \mu_\beta + \eta_i \le 0 \\ 1\left[(\alpha + c_0) + \ln(e^{\mu_\beta + \eta_i} - 1) + c_1 z + \epsilon_i + \xi_i > 0\right] & \text{if } \mu_\beta + \eta_i > 0 \end{cases} \end{aligned}$$

For estimation, we reparameterize Σ by Cholesky decomposition as:

$$\Sigma = LL', \quad \text{where} \quad L = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix},$$

Note that "L = chol(Sigma,'lower')" gives the corresponding lower triangular matrix in Matlab. Then, we have

$$\begin{pmatrix} \eta \\ \epsilon \\ \xi \end{pmatrix} = Lu, \quad \text{where} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \overset{iid}{\sim} N(0, I).$$

We set $\lambda_{33} = 1$ for identification and estimate $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{21}, \lambda_{22}, \lambda_{31}, \lambda_{32})'$ instead of $(\sigma_\xi, \sigma_\epsilon, \sigma_\eta, \rho_{\xi\epsilon}, \rho_{\xi\eta}, \rho_{\epsilon\eta})'$. With this reparametrization, we may write

$$\eta = \lambda_{11}u_1, \quad \epsilon = \lambda_{21}u_1 + \lambda_{22}u_2, \quad \xi = \lambda_{31}u_1 + \lambda_{32}u_2 + u_3.$$

Suppose that $z \in \{0, 1\}$ represents an instrument that shifts a fixed cost of importing.

We observe the data set $\{\ln \pi_i, d_i, z_i\}_{i=1}^n$, where π_i is the i-th firm's profit. We are interested in estimating the model parameter $\theta = (\alpha, c_0, c_1, \mu_\beta, \boldsymbol{\lambda}')'$.

1. We first consider the case when β is not random and given by some fixed value $\beta > 0$ and $\lambda_{11} = \lambda_{21} = \lambda_{31} = 0$. In this case, we have $\epsilon = \lambda_{22}u_2$ and $\xi = \lambda_{32}u_2 + u_3$. Define $\epsilon_i(\theta) := \ln \pi_i - \alpha - \beta d_i$ and $u_{2i}(\theta) = \epsilon_i(\theta)/\lambda_{22}$. We estimate $\theta = (\alpha, c_0, c_1, \beta, \lambda_{22}, \lambda_{32})'$ with $\lambda_{33} = 1$.

(a) Show that the likelihood function of the i-th observation is given by

$$L_i(\theta) = \frac{1}{\lambda_{22}}\phi\left(\frac{\epsilon_i(\theta)}{\lambda_{22}}\right)\Pr(d_i = 1|\epsilon_i(\theta);\theta)^{d_i}\left[1 - \Pr(d_i = 1|\epsilon_i(\theta);\theta)\right]^{1-d_i}, \text{ where}$$

$$\begin{aligned}\Pr(d_i = 1|\epsilon_i(\theta);\theta) &= \Pr\Big((\alpha + c_0) + \ln(e^\beta - 1) + c_1 z + \epsilon_i(\theta) + \xi_i > 0|\epsilon_i(\theta)\Big) \\ &= \Pr\Big((\alpha + c_0) + \ln(e^\beta - 1) + c_1 z + \epsilon_i(\theta) + \lambda_{32}u_{2i}(\theta) > -u_{3i}|\epsilon_i(\theta)\Big) \\ &= \Phi\Big((\alpha + c_0) + \ln(e^\beta - 1) + c_1 z + \epsilon_i(\theta) + \lambda_{32}u_{2i}(\theta)\Big).\end{aligned}$$

The maximum likelihood estimator is given by $\hat{\theta} = \arg\max_\theta \sum_{i=1}^n \ln L_i(\theta)$.

(b) Generate a data set with $n = 10000$ by randomly sampling from the model under $(\alpha, \beta, c_0, c_1) = (0, 0.5, -1, 2)$, $\Pr(z_i = 0)$=$\Pr(z_i = 1) = 0.5$, $(\lambda_{22}, \lambda_{32}, \lambda_{33}) = (1, 0, 1)$ and $\lambda_{11} = \lambda_{12} = \lambda_{13} = 0$ (i.e., β is constant).

i. Estimate β by OLS and IV using z_i as an instrument for d_i, and report the standard errors.

ii. Estimate θ by maximum likelihood, and report the standard errors. Note we set $\lambda_{33} = 1$ for identification and estimate $\theta = (\alpha, \beta, c_0, c_1, \lambda_{22}, \lambda_{32})'$. Discuss weather the MLE is consistent or not.

2. Now, we consider the case when β is random and correlated with ϵ and ξ. Define

$$\epsilon_i(\theta, u_1) := \ln \pi_i - \alpha - (\mu_\beta + \lambda_{11}u_1)d_i \quad \text{and} \quad u_{2i}(\theta, u_1) = (\epsilon_i(\theta, u_1) - \lambda_{21}u_1)/\lambda_{22},$$

where the latter follows from $\epsilon = \lambda_{21}u_1 + \lambda_{22}u_2$.

(a) Show that the likelihood function of the i-th observation *conditional on the value of* u_1

is given by

$$L_i(\theta, u_1) = \frac{1}{\lambda_{22}} \phi\left(u_{2i}(\theta, u_1)\right) \Pr(d_i = 1|\theta, u_1)^{d_i} \left[1 - \Pr(d_i = 1|\theta, u_1)\right]^{1-d_i}, \text{ where}$$

$$\Pr(d_i = 1|\theta, u_3) = \begin{cases} 0 & \text{if } \mu_\beta + \lambda_{11}u_1 \leq 0, \\ \Phi\left((\alpha + c_0) + \ln(e^{\mu_\beta + \lambda_{11}u_1} - 1) + c_1 z + \epsilon_i(\theta, u_1) + \lambda_{31}u_1 + \lambda_{32}u_{2i}(\theta, u_1)\right) & \text{if } \mu_\beta + \lambda_{11}u_1 > 0. \end{cases}$$

so that we can compute the likelihood function of the i-th observation by intergrating out u_1 from $L_i(\theta, u_1)$ as

$$L_i(\theta) = \int L_i(\theta, u_1)\phi(u_1)du_1$$

and the maximum likelihood estimator is given by $\hat{\theta} = \arg\max_\theta \sum_{i=1}^n \ln L_i(\theta)$. We may approximate the integral $\int L_i(\theta, u_1)\phi(u_1)du_1$ by using simulation.

(b) Generate a data set with $n = 10000$ by randomly sampling from the model under $(\alpha, \mu_\beta, c_0, c_1) = (0, 0.5, -1, 3)$, $\Pr(z_i = 0)$=$\Pr(z_i = 1) = 0.5$, $(\lambda_{11}, \lambda_{21}, \lambda_{22}, \lambda_{31}, \lambda_{32}) = (1.0, -0.8, 0.6, -0.26, -0.78)$ and we set $\lambda_{33} = 1$.

i. Compute the implied value of $(\sigma_\xi, \sigma_\epsilon, \sigma_\eta, \rho_{\xi\epsilon}, \rho_{\xi\eta}, \rho_{\epsilon\eta})'$ for this model.

ii. Estimate β by OLS and IV using z_i as an instrument for d_i, and report the standard errors. Discuss if it is possible to have that the IV estimate is larger than the OLS estimate for the model with hetergenous effects.

iii. Estimate θ by maximum likelihood, and report the standard errors.

iv. Generate the artificial data with $n = 10000$ given the estimated coefficient of θ and compute the average value of β_i for those who choose to import (i.e., $d_i = 1$) and the average value of β_i for those choose not to import (i.e., $d_i = 0$). Explain why the TT is larger than the TUT.

2 Discontinuity as an instrument

Consider the estimation of treatment effects with regression discontinuity design. Let $Y(0)$ and $Y(1)$ be two potential outcomes. We observe $Y(0)$ if $W = 0$ and $Y(1)$ if $W = 1$, where $W \in \{0, 1\}$ denotes the treatment variable, so that $Y_i = (1 - W_i)Y_i(0) + W_i Y_i(1)$. The model for the observed outcome can be written as $Y_i = \alpha_0 + \beta_i W_i + \epsilon_i$, where α_0 is a constant so that $Y_i(0) = \alpha_0 + \epsilon_i$ and $Y_i(1) - Y_i(0) = \beta_i$. Let $W_i(z)$ be a random function of z. Let Z_i be an observed random variable such that $E[W_i|Z_i = z] = \Pr(W_i = 1|Z_i = z)$ is discontinuous at z_0.

We assume that

(RD). (i) The limits $W^+ := \lim_{z \downarrow z_0^+} E[W_i|Z_i = z]$ and $W^- := \lim_{z \uparrow z_0^-} E[W_i|Z_i = z]$ exist. (ii) $W^+ \neq W^-$.

(A1). $E[\epsilon_i|Z_i = z]$ is continuous in z at z_0.

(A3). (i) $(\beta_i, W_i(z))$ is jointly independent of Z_i when z is in the neighbourhood of z_0. (ii) There exists $\delta > 0$ such that $W_i(z_0 + e) \geq W_i(z_0 - e)$ with probability one for all $0 < e < \delta$.

Prove that

$$\lim_{e \downarrow 0^+} E[\beta_i|W_i(z_0 + e) - W_i(z_0 - e) = 1] = \frac{\lim_{z \downarrow z_0^+} E[Y_i|Z_i = z] - \lim_{z \uparrow z_0^-} E[Y_i|Z_i = z]}{\lim_{z \downarrow z_0^+} E[W_i|Z_i = z] - \lim_{z \uparrow z_0^-} E[W_i|Z_i = z]}.$$

Please be specific about which assumptions are used for each line of your proof.

Hint: A set of assumptions for LATE for $Z_i \in \{0, 1\}$ is given by:

(LATE-A1). $(Y_i(0), Y_i(1), W_i(0), W_i(1))$ is independent of Z_i,

(LATE-A2). Y_i is a function of only $W_i(Z_i)$ and not Z_i directly,

(LATE-A3). $E[W_i|Z_i]$ is a nondegenerate function of Z_i,

(LATE-A4). $W_i(1) \geq W_i(0)$ for all i.

(LATE-A3) corresponds to Assumption (RD). (LATE-A2) corresponds to Assumption (A1). (LATE-A1) corresponds to Assumption (A3)(i). (LATE-A4) corresponds to Assumption (A3)(ii). Write a proof for LATE under (LATE-A1)–(LATE-A4) and repeat the same line of argument for the regression discontinuity design to show

$$\frac{E[Y_i|Z_i = z_0 + e] - E[Y_i|Z_i = z_0 - e]}{E[W_i|Z_i = z_0 + e] - E[W_i|Z_i = z_0 - e]} = E[\beta_i|W_i(z_0 + e) - W_i(z_0 - e) = 1].$$